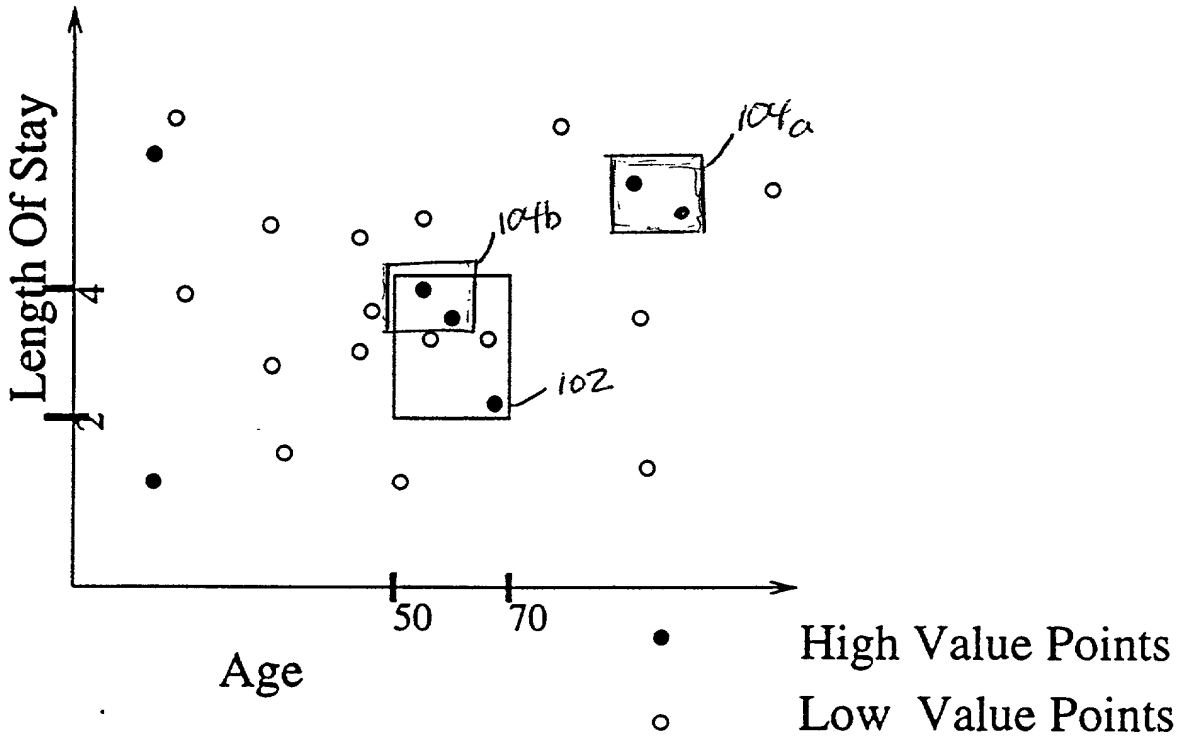


Hyper Box defined as 50 < Age < 70 and 2 <= Stay Length <= 4

FIG. 1

Doctor ID	A	B	C	D
Expected cost	150	200	900	1000

Doctor ID Histogram

$$\text{Expected cost} = \frac{\sum \text{cost}}{\#\text{ of instances of the discrete attribute value}}$$

Index	Doctor ID	Age	Length of Stay	Cost
	A	20	1	100
2	A	40	2	300
3	B	16	1	100
4	B	60	3	1000
5	C	70	2	900
⋮	⋮	⋮	⋮	⋮
100	D	50	4	1500

Relational Table of Data

FIG. 2

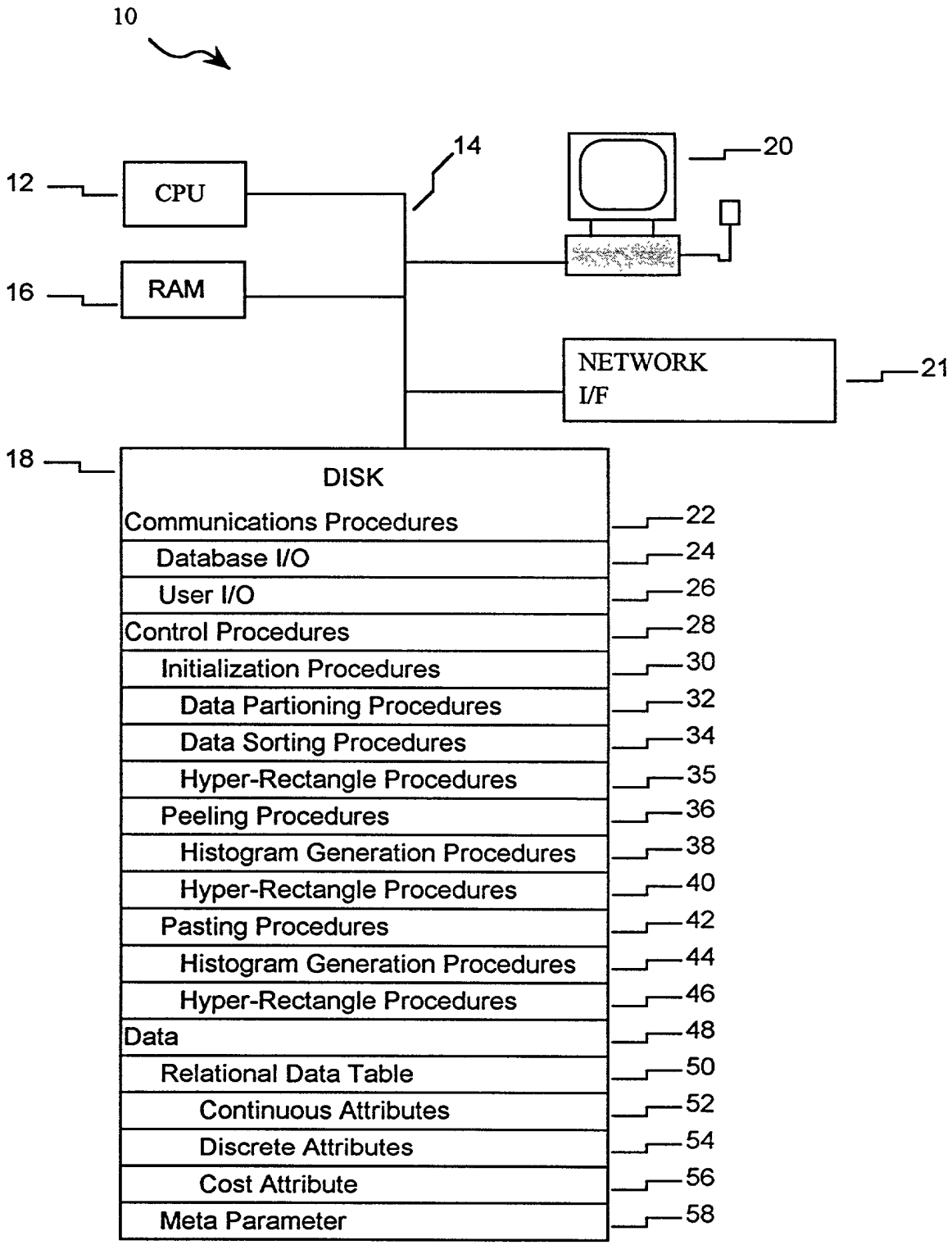

10
12
CPU
14
20
16
RAM
NETWORK
I/F
21
18
DISK
Communications Procedures
22
Database I/O
24
User I/O
26
Control Procedures
28
Initialization Procedures
30
Data Partioning Procedures
32
Data Sorting Procedures
34
Hyper-Rectangle Procedures
35
Peeling Procedures
36
Histogram Generation Procedures
38
Hyper-Rectangle Procedures
40
Pasting Procedures
42
Histogram Generation Procedures
44
Hyper-Rectangle Procedures
46
Data
48
Relational Data Table
50
Continuous Attributes
52
Discrete Attributes
54
Cost Attribute
56
Meta Parameter
58

FIG. 3

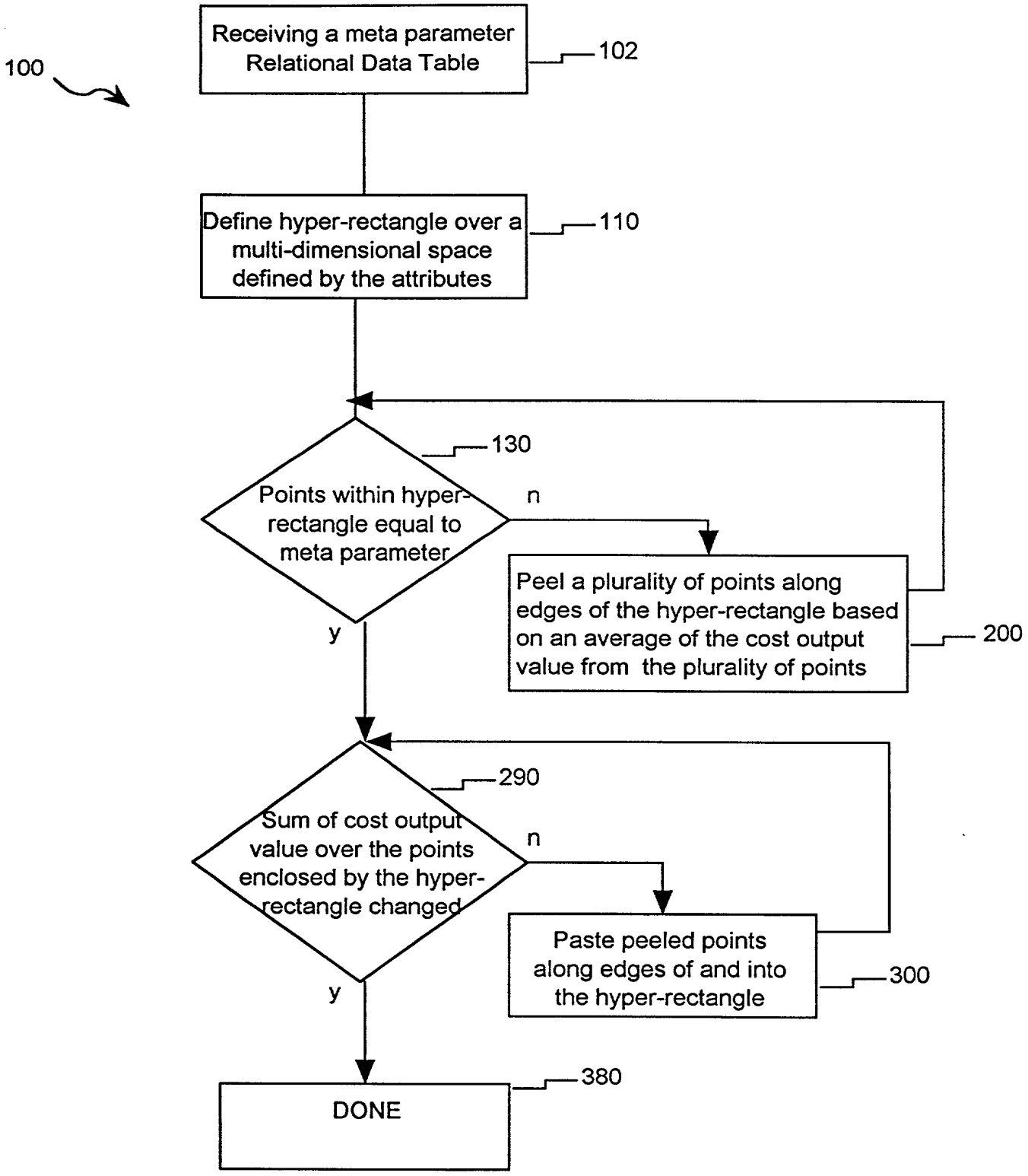

100
Receiving a meta parameter Relational Data Table
102
Define hyper-rectangle over a multi-dimensional space defined by the attributes
110
130
Points within hyper-rectangle equal to meta parameter
n
y
Peel a plurality of points along edges of the hyper-rectangle based on an average of the cost output value from the plurality of points
200
290
Sum of cost output value over the points enclosed by the hyper-rectangle changed
n
y
Paste peeled points along edges of and into the hyper-rectangle
300
DONE
380

FIG. 4

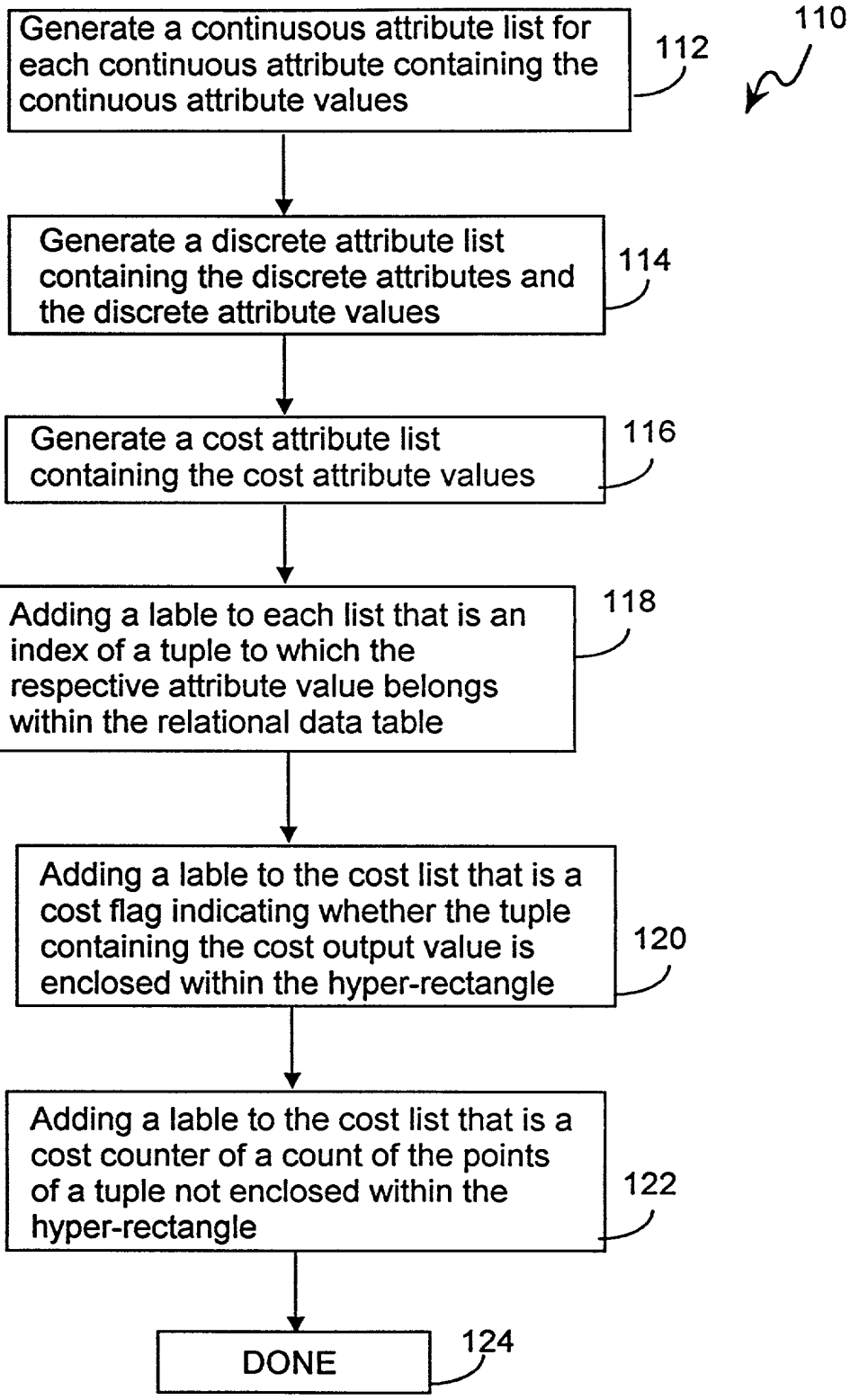

Generate a continusous attribute list for each continuous attribute containing the continuous attribute values
112
110
Generate a discrete attribute list containing the discrete attributes and the discrete attribute values
114
Generate a cost attribute list containing the cost attribute values
116
Adding a lable to each list that is an index of a tuple to which the respective attribute value belongs within the relational data table
118
Adding a lable to the cost list that is a cost flag indicating whether the tuple containing the cost output value is enclosed within the hyper-rectangle
120
Adding a lable to the cost list that is a cost counter of a count of the points of a tuple not enclosed within the hyper-rectangle
122
DONE
124

FIG. 5

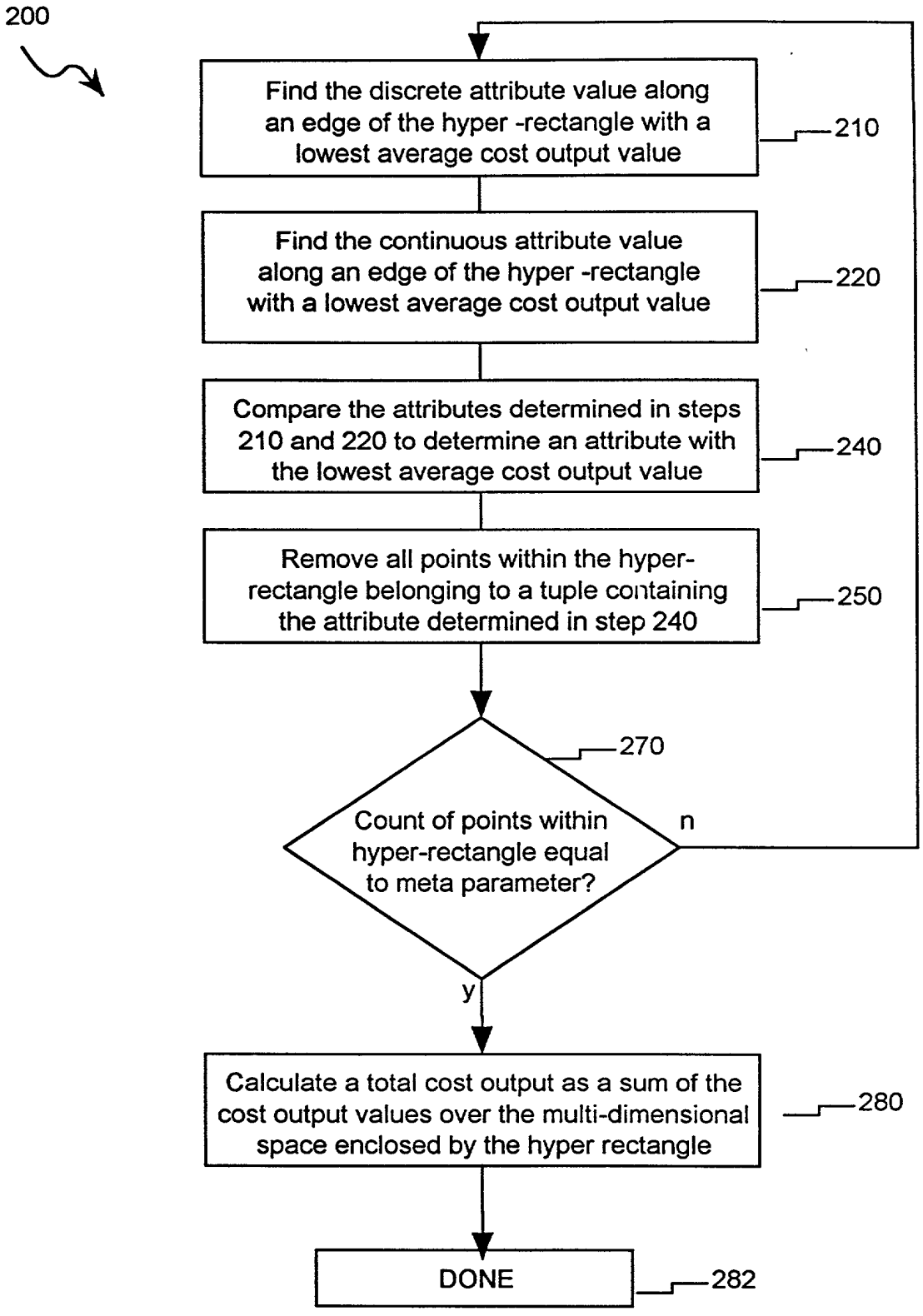

200
Find the discrete attribute value along an edge of the hyper -rectangle with a lowest average cost output value
210
Find the continuous attribute value along an edge of the hyper -rectangle with a lowest average cost output value
220
Compare the attributes determined in steps 210 and 220 to determine an attribute with the lowest average cost output value
240
Remove all points within the hyper-rectangle belonging to a tuple containing the attribute determined in step 240
250
270
Count of points within hyper-rectangle equal to meta parameter?
n
y
Calculate a total cost output as a sum of the cost output values over the multi-dimensional space enclosed by the hyper rectangle
280
DONE
282

FIG. 6

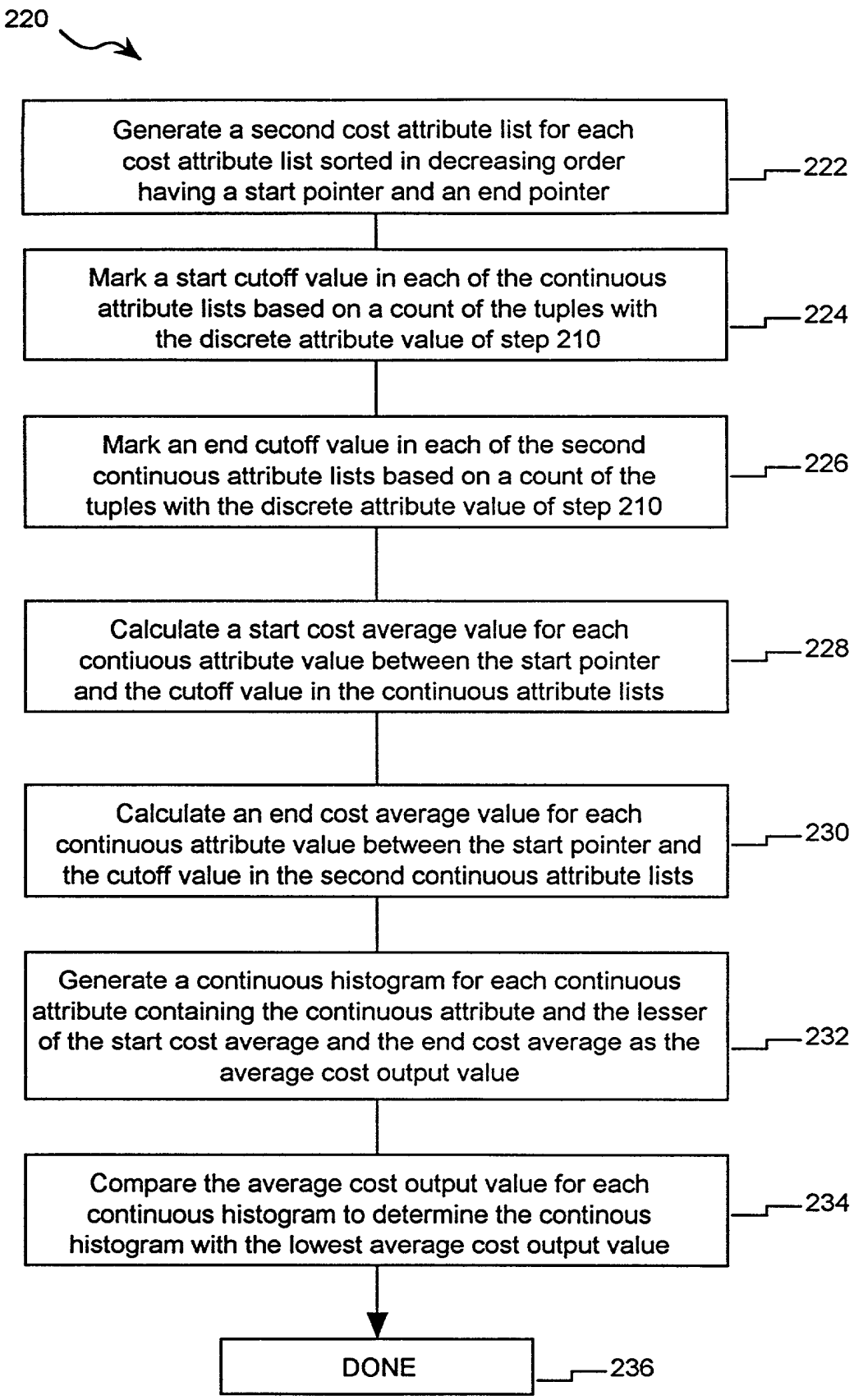

220
Generate a second cost attribute list for each cost attribute list sorted in decreasing order having a start pointer and an end pointer
222
Mark a start cutoff value in each of the continuous attribute lists based on a count of the tuples with the discrete attribute value of step 210
224
Mark an end cutoff value in each of the second continuous attribute lists based on a count of the tuples with the discrete attribute value of step 210
226
Calculate a start cost average value for each contiuous attribute value between the start pointer and the cutoff value in the continuous attribute lists
228
Calculate an end cost average value for each continuous attribute value between the start pointer and the cutoff value in the second continuous attribute lists
230
Generate a continuous histogram for each continuous attribute containing the continuous attribute and the lesser of the start cost average and the end cost average as the average cost output value
232
Compare the average cost output value for each continuous histogram to determine the continous histogram with the lowest average cost output value
234
DONE
236

FIG. 7

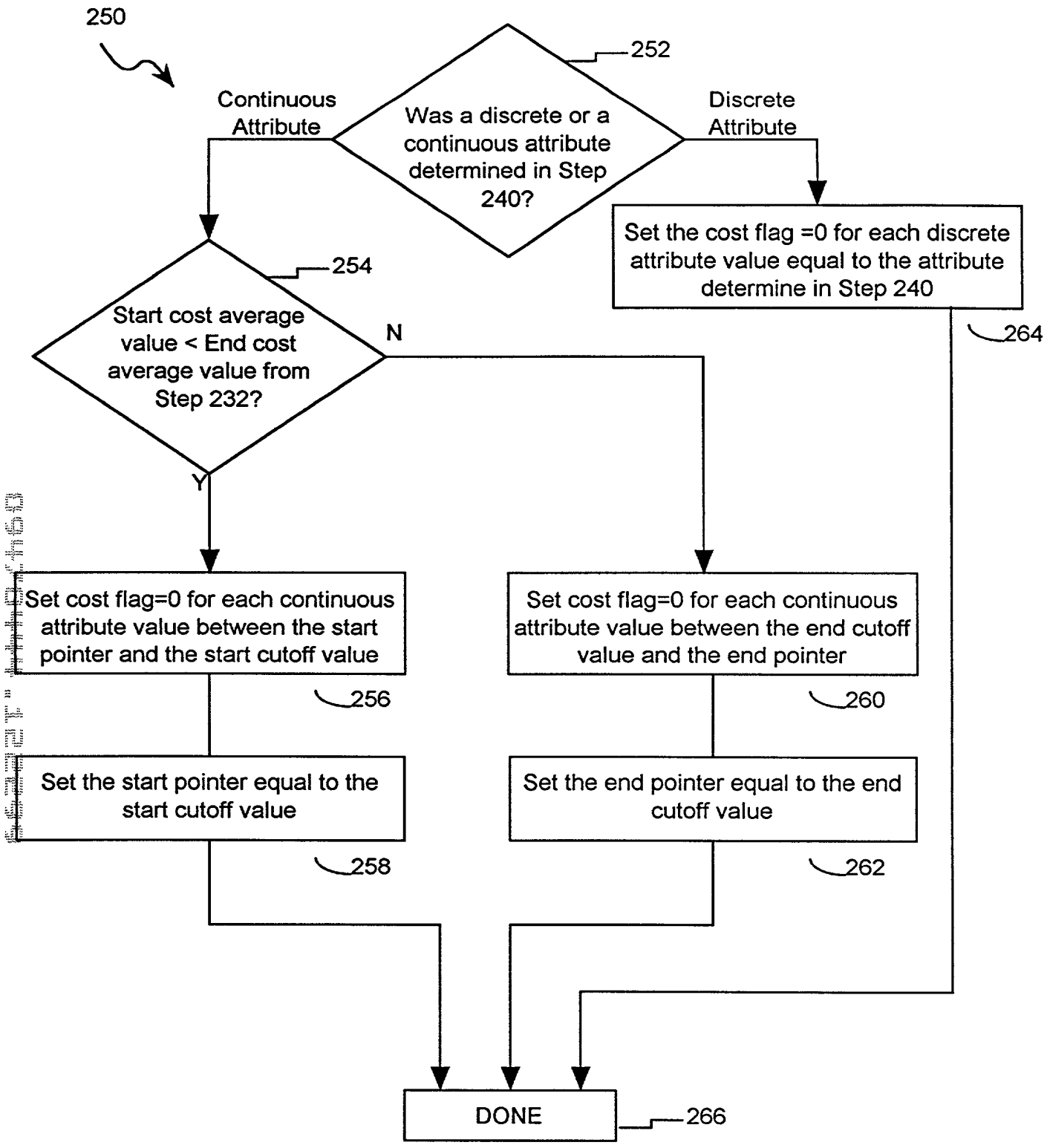

250
252
Was a discrete or a continuous attribute determined in Step 240?
Continuous Attribute
Discrete Attribute
Set the cost flag =0 for each discrete attribute value equal to the attribute determine in Step 240
264
254
Start cost average value < End cost average value from Step 232?
N
Y
Set cost flag=0 for each continuous attribute value between the start pointer and the start cutoff value
256
Set the start pointer equal to the start cutoff value
258
Set cost flag=0 for each continuous attribute value between the end cutoff value and the end pointer
260
Set the end pointer equal to the end cutoff value
262
DONE
266

FIG. 8

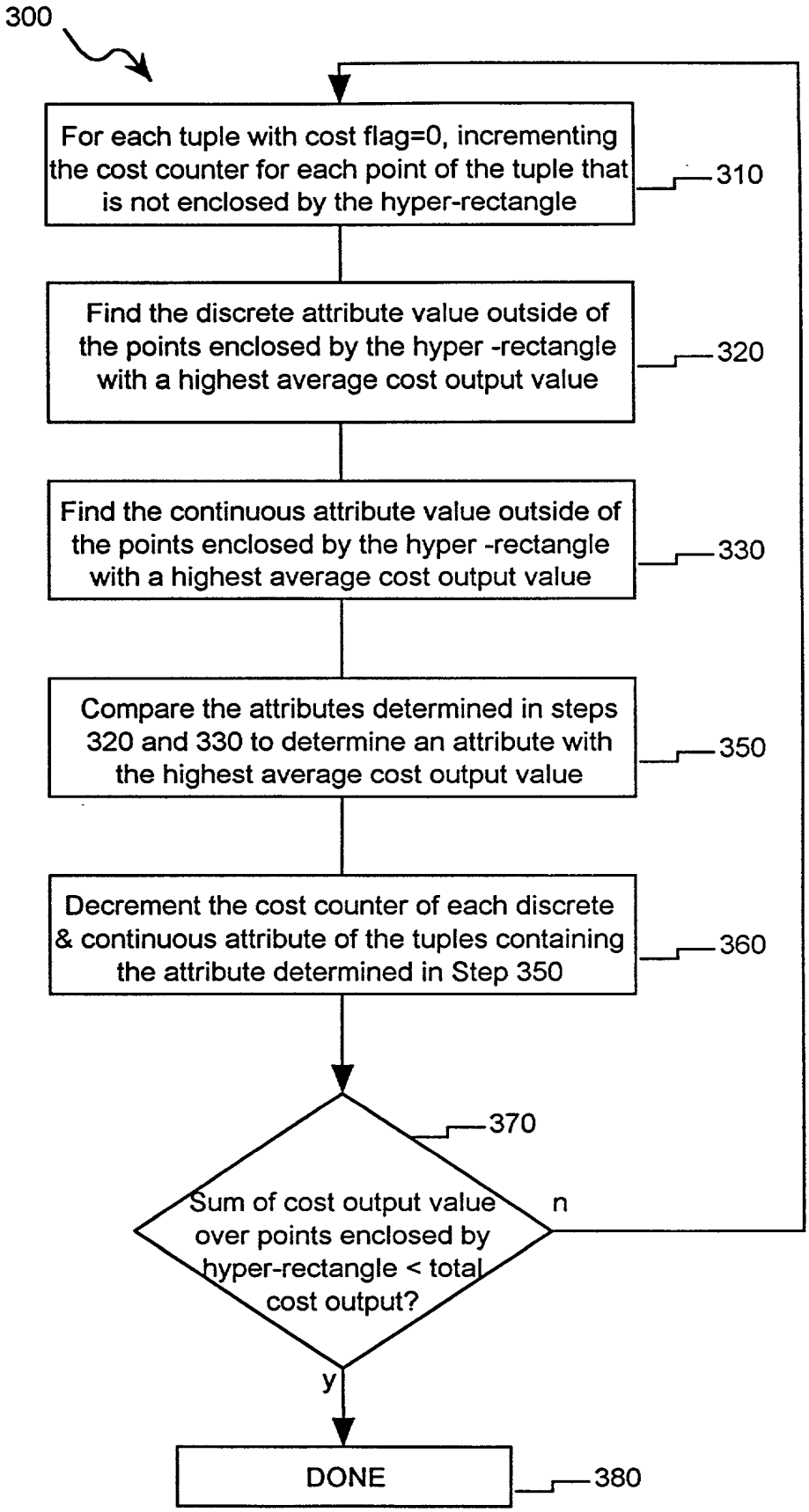

300
For each tuple with cost flag=0, incrementing the cost counter for each point of the tuple that is not enclosed by the hyper-rectangle
310
Find the discrete attribute value outside of the points enclosed by the hyper -rectangle with a highest average cost output value
320
Find the continuous attribute value outside of the points enclosed by the hyper -rectangle with a highest average cost output value
330
Compare the attributes determined in steps 320 and 330 to determine an attribute with the highest average cost output value
350
Decrement the cost counter of each discrete & continuous attribute of the tuples containing the attribute determined in Step 350
360
370
Sum of cost output value over points enclosed by hyper-rectangle < total cost output?
n
y
DONE
380

FIG. 9

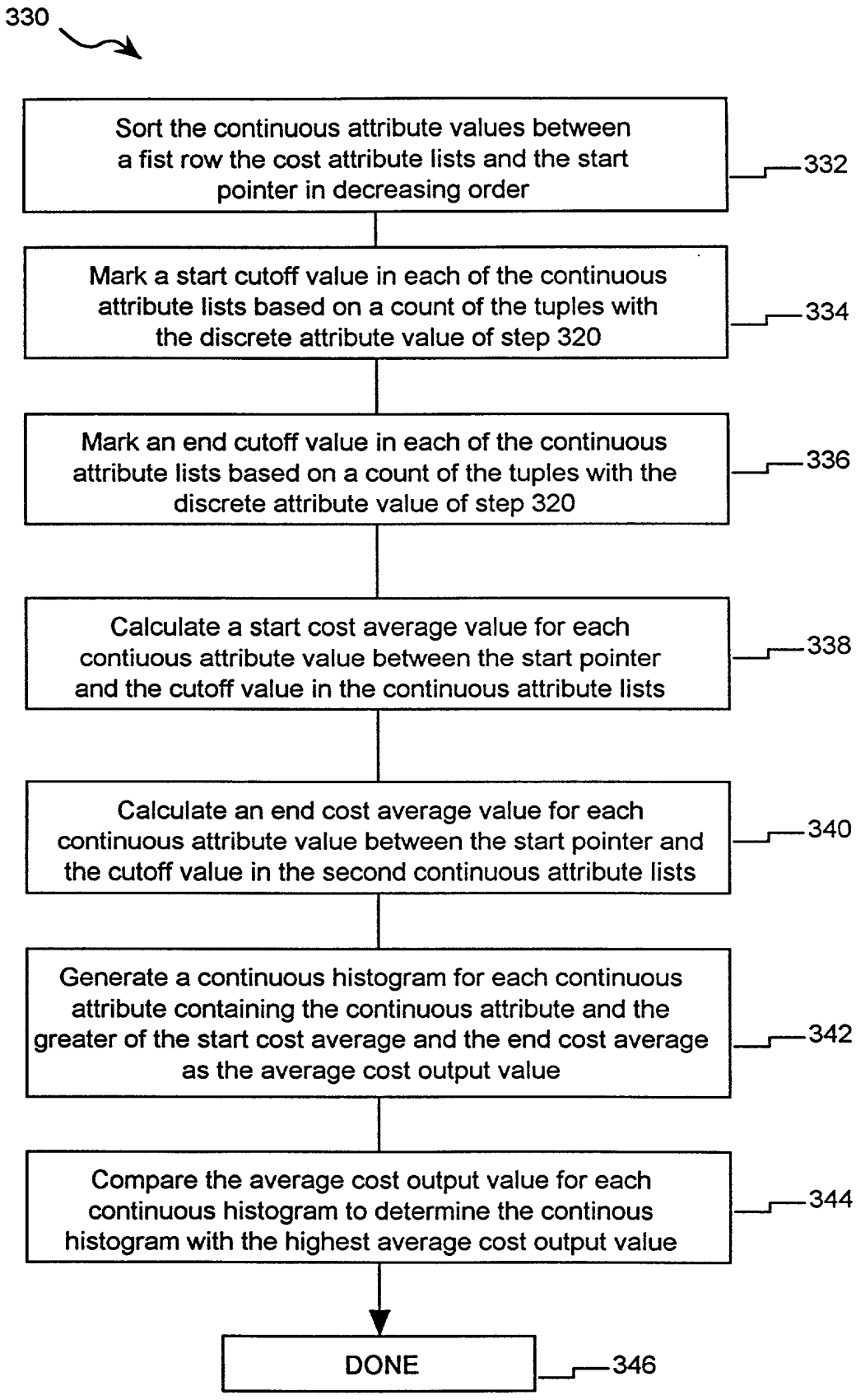

330
Sort the continuous attribute values between a fist row the cost attribute lists and the start pointer in decreasing order
332
Mark a start cutoff value in each of the continuous attribute lists based on a count of the tuples with the discrete attribute value of step 320
334
Mark an end cutoff value in each of the continuous attribute lists based on a count of the tuples with the discrete attribute value of step 320
336
Calculate a start cost average value for each contiuous attribute value between the start pointer and the cutoff value in the continuous attribute lists
338
Calculate an end cost average value for each continuous attribute value between the start pointer and the cutoff value in the second continuous attribute lists
340
Generate a continuous histogram for each continuous attribute containing the continuous attribute and the greater of the start cost average and the end cost average as the average cost output value
342
Compare the average cost output value for each continuous histogram to determine the continous histogram with the highest average cost output value
344
DONE
346

FIG. 10

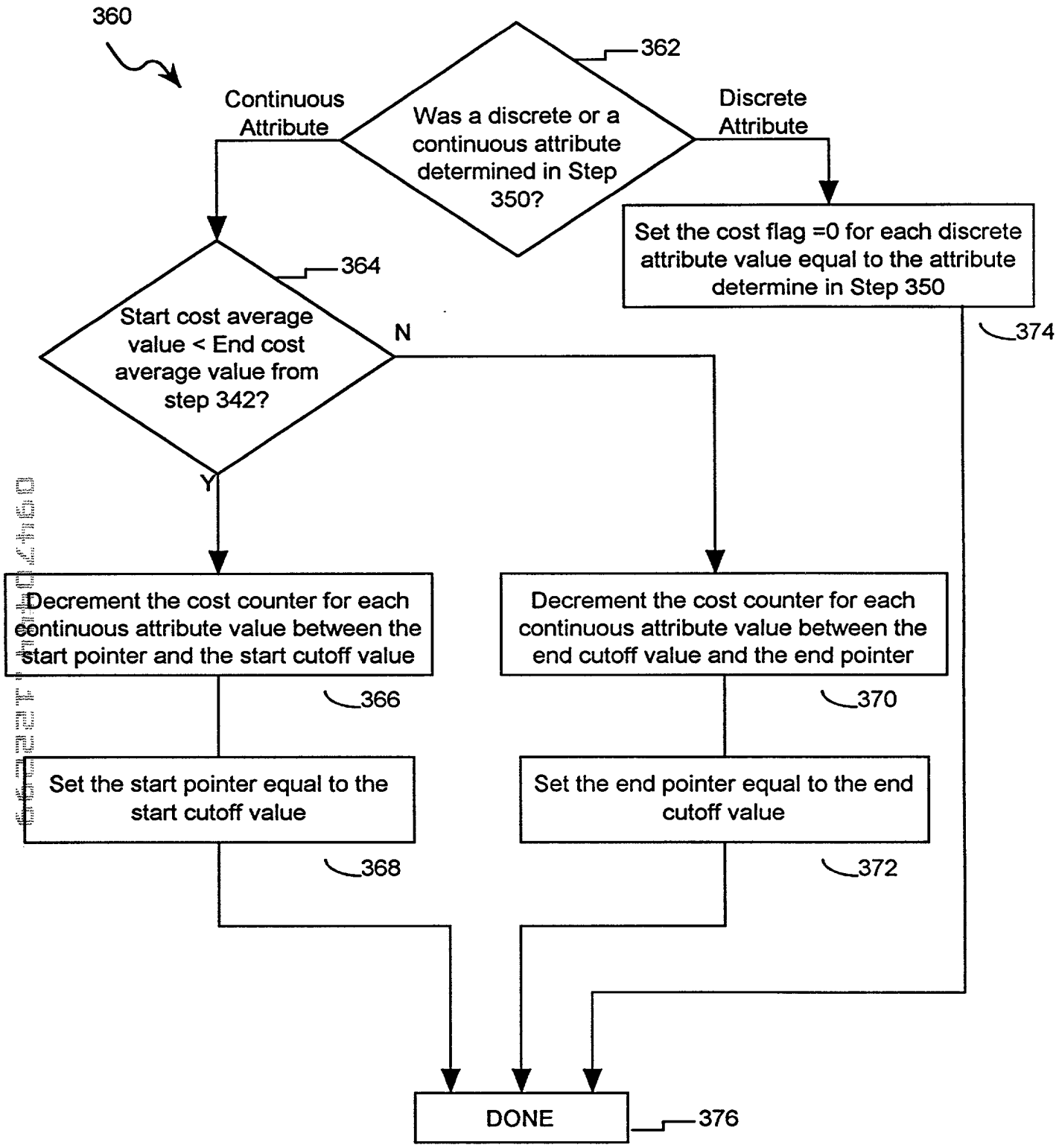

360
362
Was a discrete or a continuous attribute determined in Step 350?
Continuous Attribute
Discrete Attribute
Set the cost flag =0 for each discrete attribute value equal to the attribute determine in Step 350
374
364
Start cost average value < End cost average value from step 342?
N
Y
Decrement the cost counter for each continuous attribute value between the start pointer and the start cutoff value
366
Decrement the cost counter for each continuous attribute value between the end cutoff value and the end pointer
370
Set the start pointer equal to the start cutoff value
368
Set the end pointer equal to the end cutoff value
372
DONE
376

FIG. 11

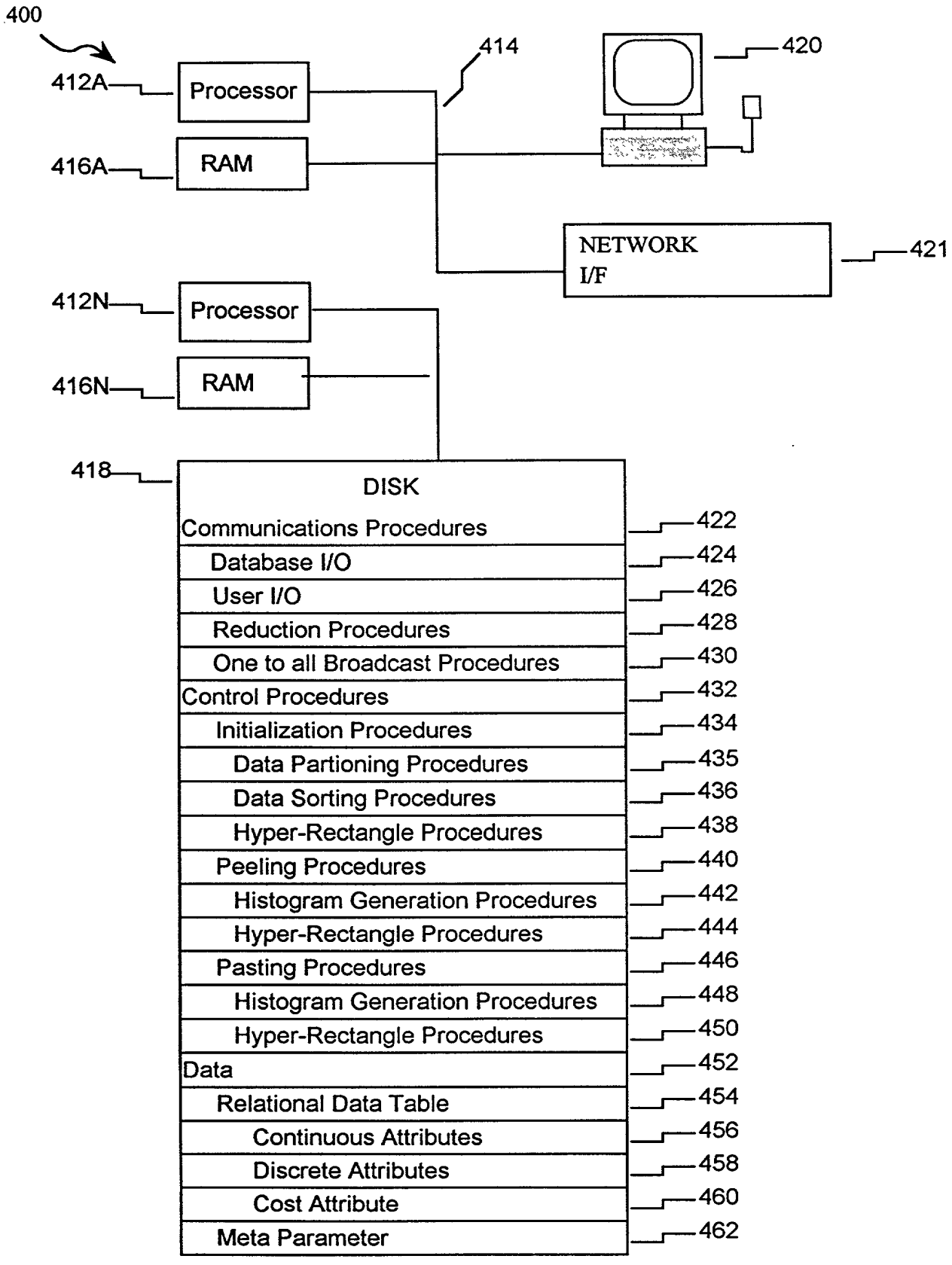

400
412A
Processor
414
420
416A
RAM
NETWORK
I/F
421
412N
Processor
416N
RAM
418
DISK
Communications Procedures
422
Database I/O
424
User I/O
426
Reduction Procedures
428
One to all Broadcast Procedures
430
Control Procedures
432
Initialization Procedures
434
Data Partioning Procedures
435
Data Sorting Procedures
436
Hyper-Rectangle Procedures
438
Peeling Procedures
440
Histogram Generation Procedures
442
Hyper-Rectangle Procedures
444
Pasting Procedures
446
Histogram Generation Procedures
448
Hyper-Rectangle Procedures
450
Data
452
Relational Data Table
454
Continuous Attributes
456
Discrete Attributes
458
Cost Attribute
460
Meta Parameter
462

FIG. 12

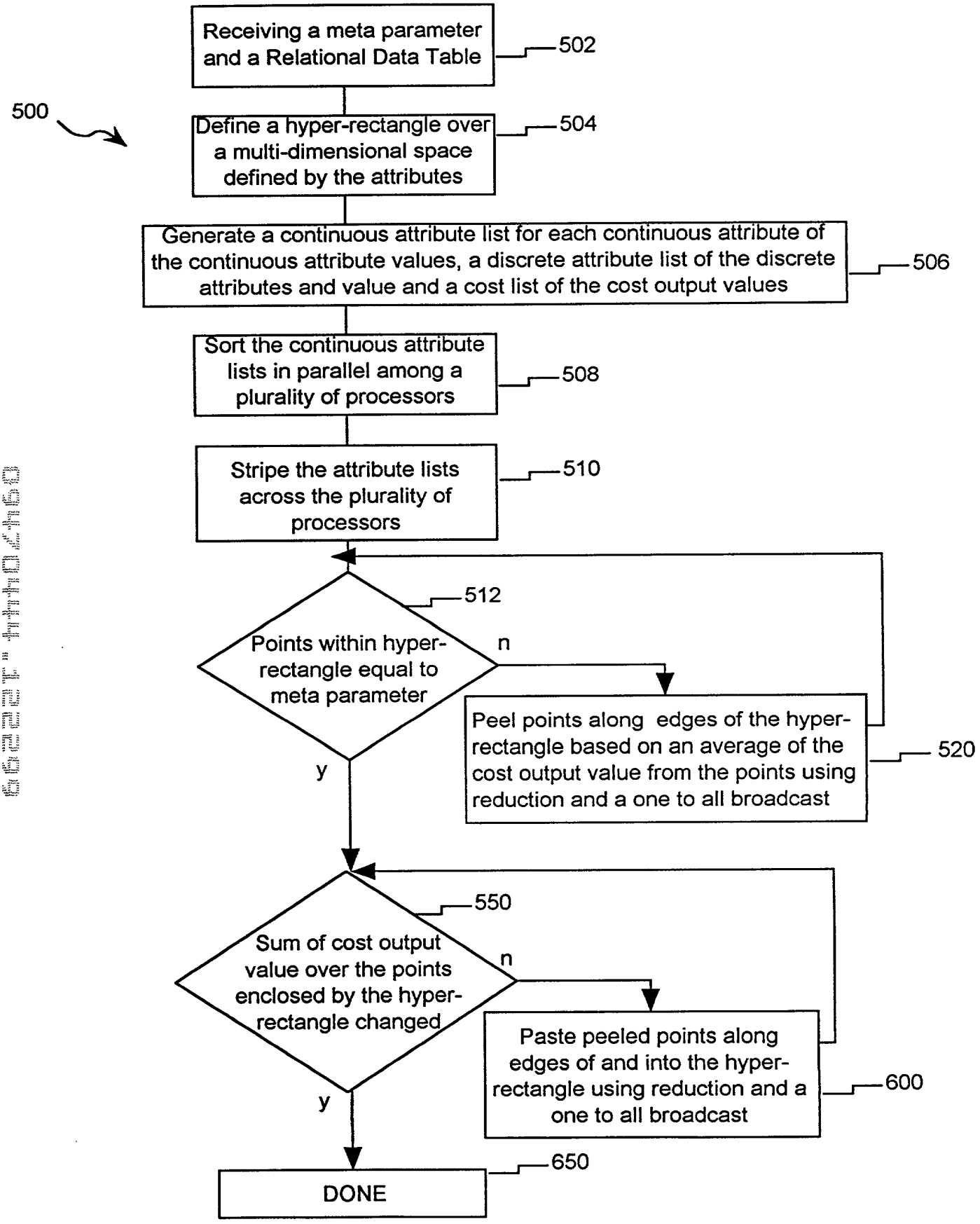
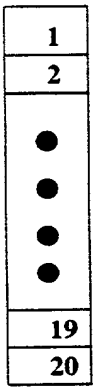
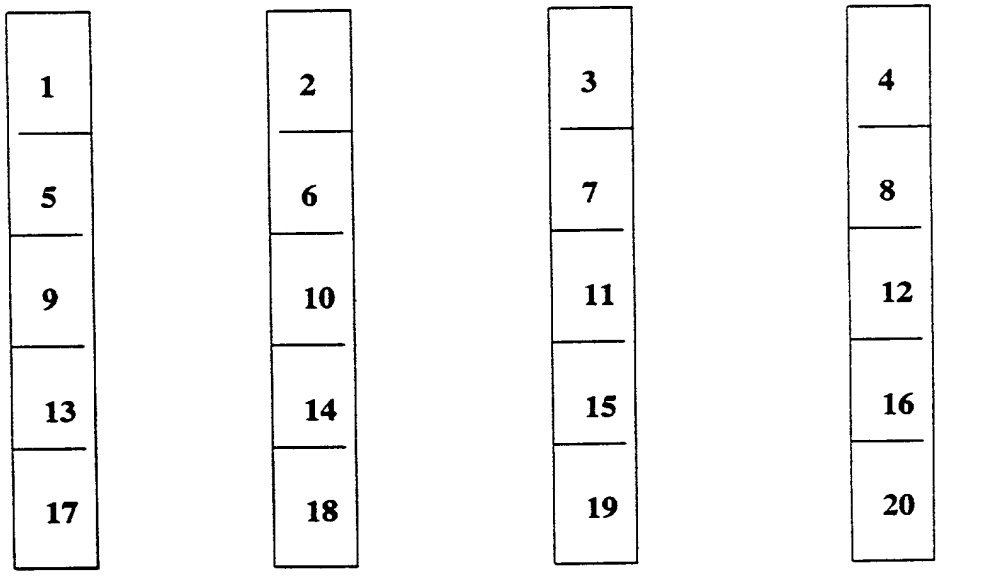

500
Receiving a meta parameter and a Relational Data Table
502
Define a hyper-rectangle over a multi-dimensional space defined by the attributes
504
Generate a continuous attribute list for each continuous attribute of the continuous attribute values, a discrete attribute list of the discrete attributes and value and a cost list of the cost output values
506
Sort the continuous attribute lists in parallel among a plurality of processors
508
Stripe the attribute lists across the plurality of processors
510
512
Points within hyper-rectangle equal to meta parameter
n
y
Peel points along edges of the hyper-rectangle based on an average of the cost output value from the points using reduction and a one to all broadcast
520
550
Sum of cost output value over the points enclosed by the hyper-rectangle changed
n
y
Paste peeled points along edges of and into the hyper-rectangle using reduction and a one to all broadcast
600
DONE
650

FIG. 13



Original sorted list



Processor 1 Processor 2 Processor 3 Processor 4

FIG. 14

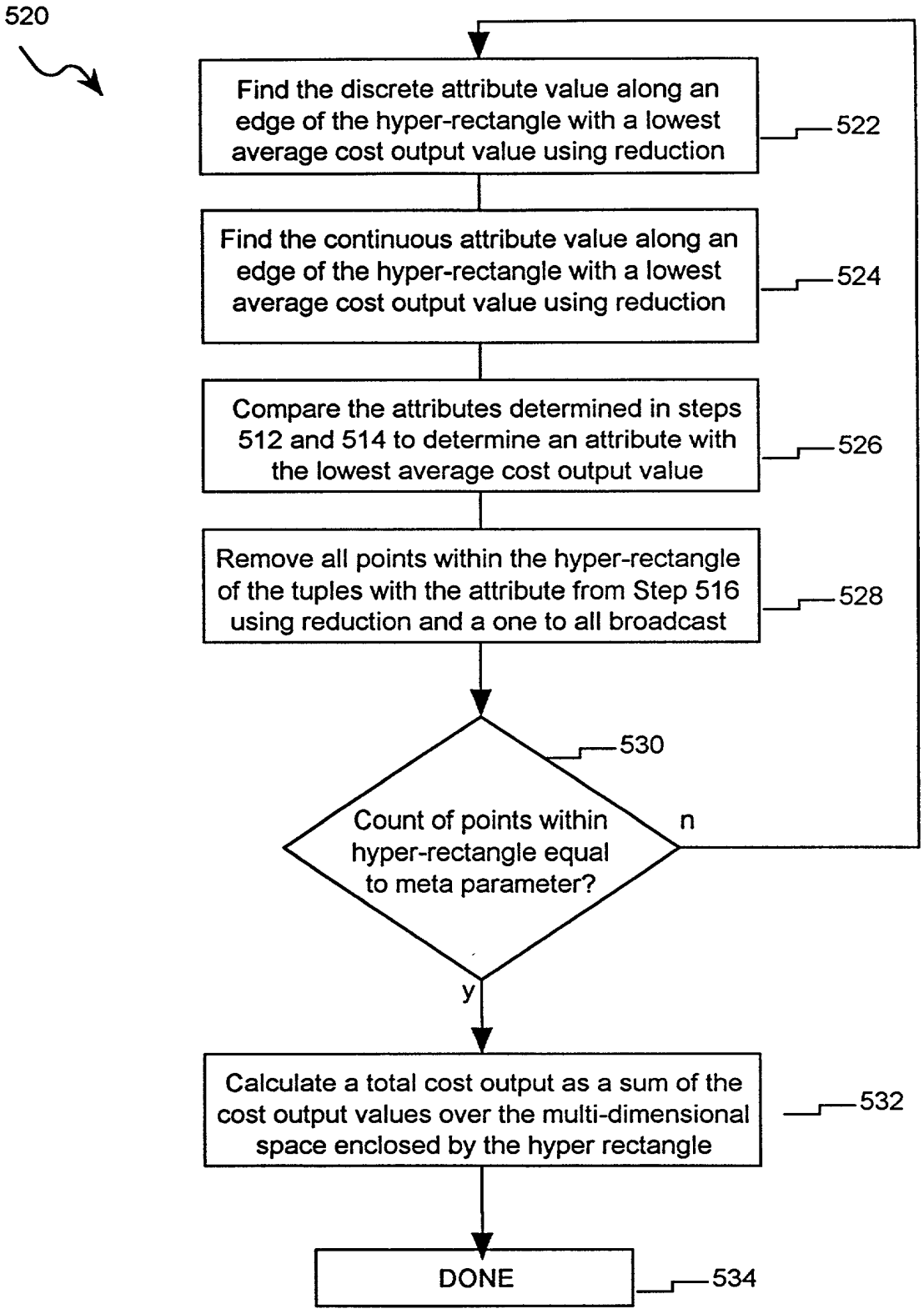

520
Find the discrete attribute value along an edge of the hyper-rectangle with a lowest average cost output value using reduction
522
Find the continuous attribute value along an edge of the hyper-rectangle with a lowest average cost output value using reduction
524
Compare the attributes determined in steps 512 and 514 to determine an attribute with the lowest average cost output value
526
Remove all points within the hyper-rectangle of the tuples with the attribute from Step 516 using reduction and a one to all broadcast
528
530
Count of points within hyper-rectangle equal to meta parameter?
n
y
Calculate a total cost output as a sum of the cost output values over the multi-dimensional space enclosed by the hyper rectangle
532
DONE
534

FIG. 15

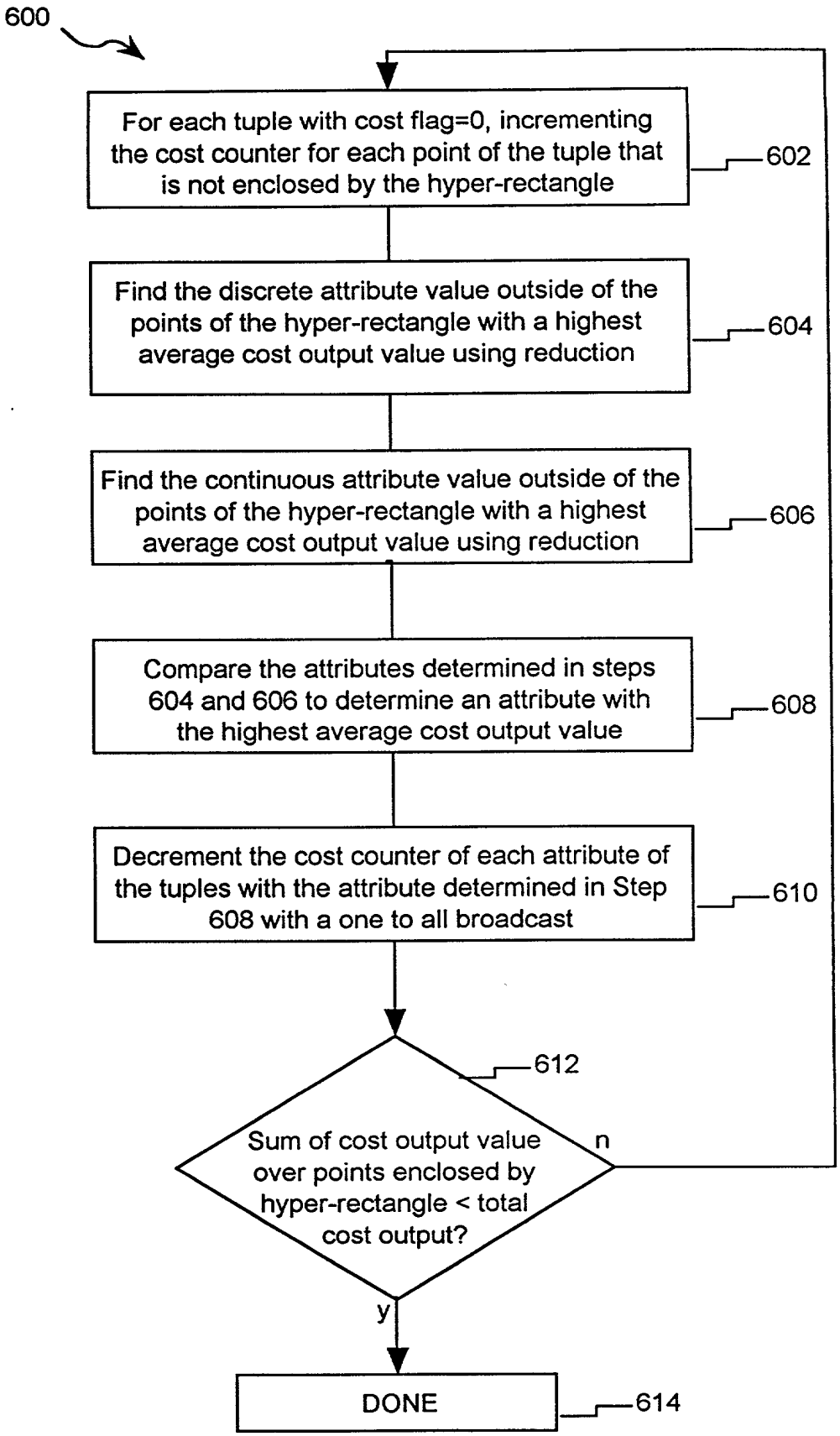

600
For each tuple with cost flag=0, incrementing the cost counter for each point of the tuple that is not enclosed by the hyper-rectangle
602
Find the discrete attribute value outside of the points of the hyper-rectangle with a highest average cost output value using reduction
604
Find the continuous attribute value outside of the points of the hyper-rectangle with a highest average cost output value using reduction
606
Compare the attributes determined in steps 604 and 606 to determine an attribute with the highest average cost output value
608
Decrement the cost counter of each attribute of the tuples with the attribute determined in Step 608 with a one to all broadcast
610
612
Sum of cost output value over points enclosed by hyper-rectangle < total cost output?
n
y
DONE
614

FIG. 16

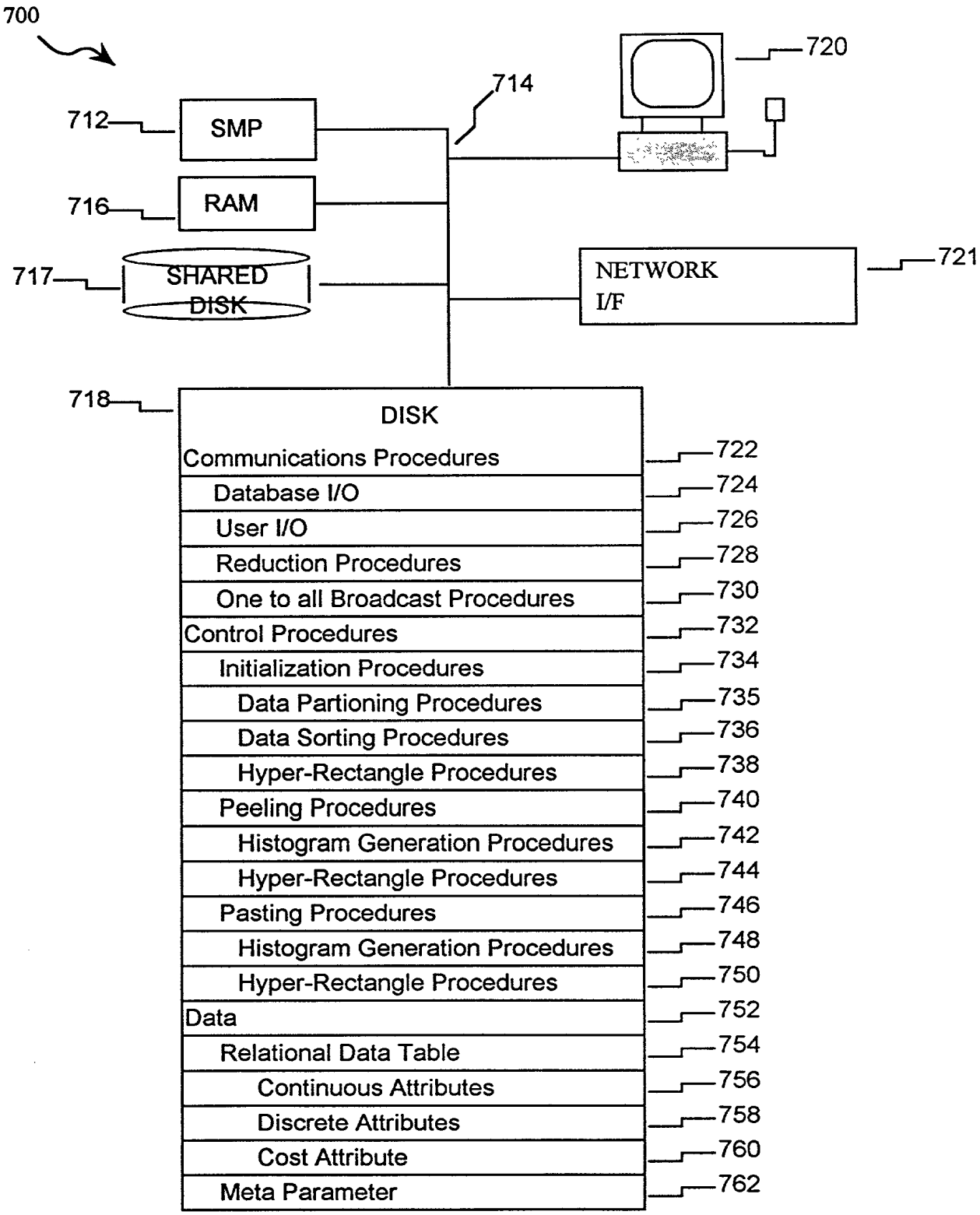

700
712
SMP
714
720
716
RAM
717
SHARED DISK
NETWORK I/F
721
718
DISK
Communications Procedures
722
Database I/O
724
User I/O
726
Reduction Procedures
728
One to all Broadcast Procedures
730
Control Procedures
732
Initialization Procedures
734
Data Partioning Procedures
735
Data Sorting Procedures
736
Hyper-Rectangle Procedures
738
Peeling Procedures
740
Histogram Generation Procedures
742
Hyper-Rectangle Procedures
744
Pasting Procedures
746
Histogram Generation Procedures
748
Hyper-Rectangle Procedures
750
Data
752
Relational Data Table
754
Continuous Attributes
756
Discrete Attributes
758
Cost Attribute
760
Meta Parameter
762

FIG. 17

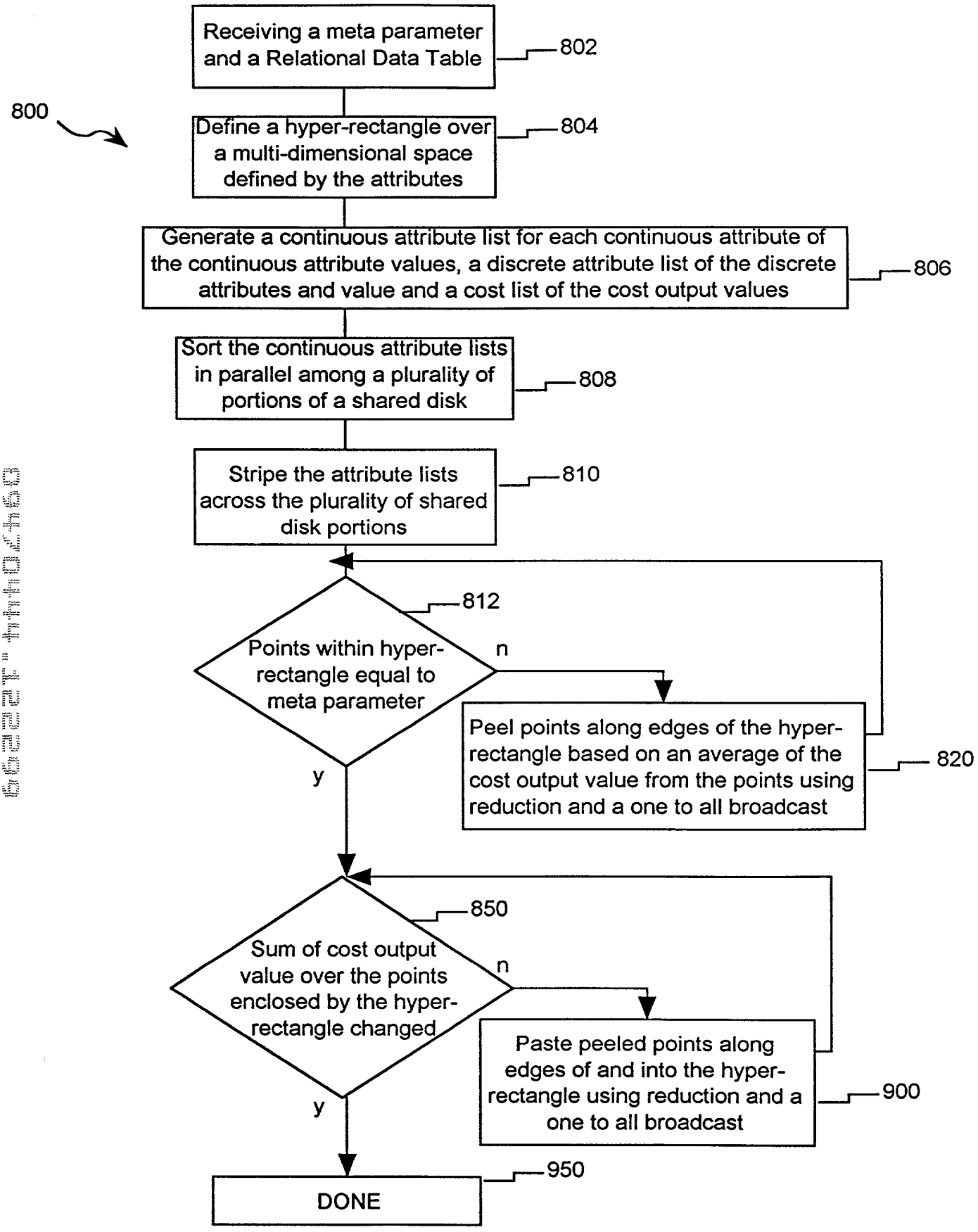

800
Receiving a meta parameter and a Relational Data Table
802
Define a hyper-rectangle over a multi-dimensional space defined by the attributes
804
Generate a continuous attribute list for each continuous attribute of the continuous attribute values, a discrete attribute list of the discrete attributes and value and a cost list of the cost output values
806
Sort the continuous attribute lists in parallel among a plurality of portions of a shared disk
808
Stripe the attribute lists across the plurality of shared disk portions
810
812
Points within hyper-rectangle equal to meta parameter
n
y
Peel points along edges of the hyper-rectangle based on an average of the cost output value from the points using reduction and a one to all broadcast
820
850
Sum of cost output value over the points enclosed by the hyper-rectangle changed
n
y
Paste peeled points along edges of and into the hyper-rectangle using reduction and a one to all broadcast
900
DONE
950

FIG. 18

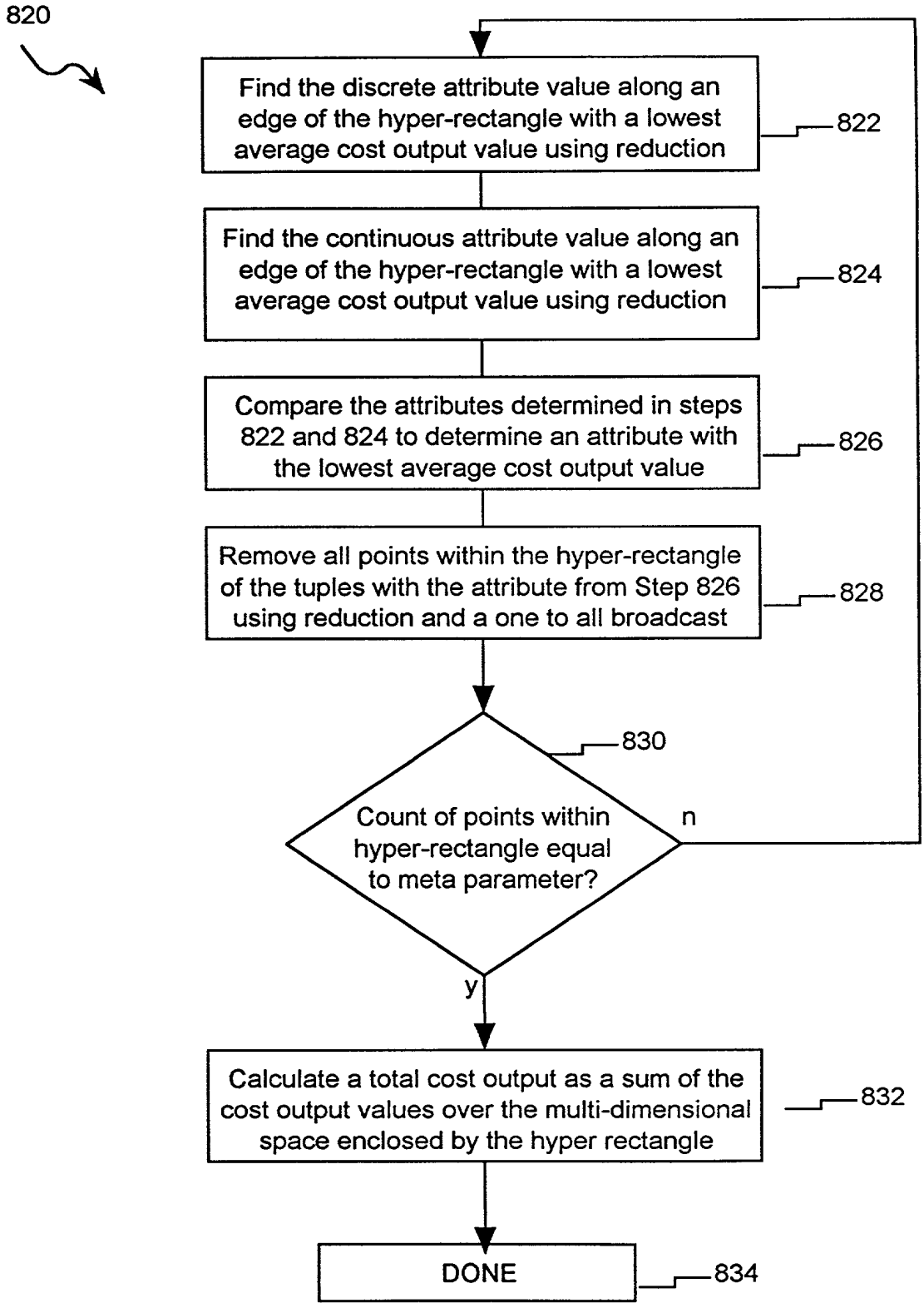

820
Find the discrete attribute value along an edge of the hyper-rectangle with a lowest average cost output value using reduction
822
Find the continuous attribute value along an edge of the hyper-rectangle with a lowest average cost output value using reduction
824
Compare the attributes determined in steps 822 and 824 to determine an attribute with the lowest average cost output value
826
Remove all points within the hyper-rectangle of the tuples with the attribute from Step 826 using reduction and a one to all broadcast
828
830
Count of points within hyper-rectangle equal to meta parameter?
n
y
Calculate a total cost output as a sum of the cost output values over the multi-dimensional space enclosed by the hyper rectangle
832
DONE
834

FIG. 19

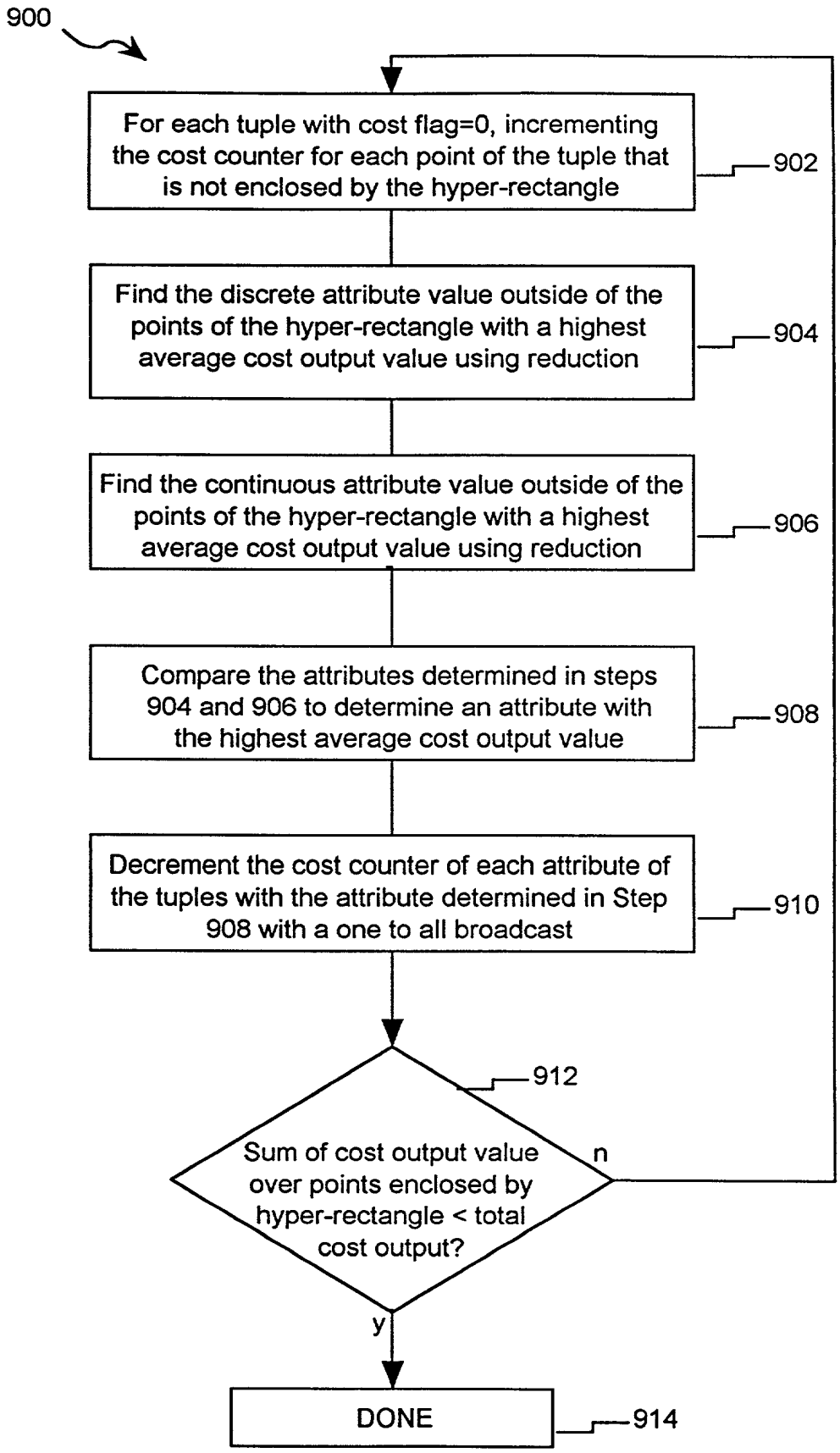

900
For each tuple with cost flag=0, incrementing the cost counter for each point of the tuple that is not enclosed by the hyper-rectangle
902
Find the discrete attribute value outside of the points of the hyper-rectangle with a highest average cost output value using reduction
904
Find the continuous attribute value outside of the points of the hyper-rectangle with a highest average cost output value using reduction
906
Compare the attributes determined in steps 904 and 906 to determine an attribute with the highest average cost output value
908
Decrement the cost counter of each attribute of the tuples with the attribute determined in Step 908 with a one to all broadcast
910
912
Sum of cost output value over points enclosed by hyper-rectangle < total cost output?
n
y
DONE
914

FIG. 20